ASYMPTOTIC ANALYSIS FOR THE GENERALIZED LANGEVIN EQUATION

M. Ottobre and G.A. Pavliotis
Department of Mathematics
Imperial College London
London SW7 2AZ, UK

March 23, 2010

Abstract

Various qualitative properties of solutions to the generalized Langevin equation (GLE) in a periodic or a confining potential are studied in this paper. We consider a class of quasi-Markovian GLEs, similar to the model that was introduced in [7]. Geometric ergodicity, a homogenization theorem (invariance principle), short time asymptotics and the white noise limit are studied. Our proofs are based on a careful analysis of a hypoelliptic operator which is the generator of an auxiliary Markov process. Systematic use of the recently developed theory of hypocoercivity [40] is made.

1 Introduction

In this paper we study various qualitative properties of solutions to the generalized Langevin equation (GLE) in $\mathbb{R}^d$

$$\ddot{q} = -\nabla V(q) - \int_0^t \gamma(t-s)\dot{q}(s)\,ds + F(t), \qquad (1)$$

where $V(q)$ is a smooth potential (confining or periodic), $F(t)$ a mean zero stationary Gaussian process with autocorrelation function $\gamma(t)$, in accordance to the fluctuation-dissipation theorem

$$\langle F(t) \otimes F(s) \rangle = \beta^{-1}\gamma(t-s)I. \qquad (2)$$

Here β stands for the inverse temperature and I for the identity matrix. The GLE equation (1), together with the fluctuation–dissipation theorem (2) appears in various applications such as surface diffusion [1] and polymer dynamics [39]. It also serves as one of the standard models of nonequilibirum statistical mechanics, describing the dynamics of a "small" Hamiltonian system (the distinguished particle) coupled to one or more heat baths which are modelled as linear wave equations with initial conditions which are distributed according to

appropriate Gibbs measures [36]. In this class of models the coupling between the distinguished particle and the heat bath is taken to be linear and is governed by a coupling function $\rho(x)$. The full Hamiltonian of the "particle+heat bath" model is

$$H(q, p, \phi, \pi) = H_{DP}(p, q) + \mathcal{H}(\phi, \pi) + \lambda q \int \rho(x) \partial_q \phi(x) \, dx \qquad (3)$$

where $H_{DP}(q, p)$ denotes the Hamiltonian of the distinguished particle whose position and momentum are denoted by q and p, respectively and $\mathcal{H}(\phi, \pi)$ is the Hamiltonian density of the wave equation where ϕ and π are the canonically conjugate field variables. The linear coupling in (3) is motivated by the dipole approximation from classical electrodynamics. By integrating out the heat bath variables and using our assumptions on the initial conditions we obtain (1), together with (2). The memory kernel $\gamma(t)$ in (1) is given by the coupling function through the formula

$$\gamma(t) = \int |\widehat{\rho}(k)|^2 e^{ikt} \, dk, \qquad (4)$$

where $\widehat{\rho}(k)$ denotes the Fourier transform of $\rho(x)$ [20, 36].

The GLE has also attracted attention in recent years in the context of mode reduction and coarse-graininig for high dimensional dynamical systems [10]. One of the models that has been studied extensively within the framework of mode elimination is the Kac-Zwanzig model [9, 41] and its variants [26, 15, 24, 3]. In this model, the heat bath is modelled as a finite dimensional system of N harmonic oscillators with random frequencies and random initial conditions distributed according to a Gibbs distribution at inverse temperature β. The heat bath can be coupled either linearly or nonlinearly with the distinguished particle [25]. Just as with model (3), we can integrate out the heat bath variables explicitly. Passing then to the thermodynamic limit $N \to +\infty$, we obtain the GLE (1). The form of the memory kernel $\gamma(t)$ depends on the choice of the distribution of the spring constants of the harmonic oscillators in the heat bath [10]. The Kac-Zwanzig model and its variants have proved to be very useful for testing various methodologies and techniques such as transition state theory [2, 16].

The GLE (1) is a stochastic integrodifferential equation which is equivalent to the original infinite dimensional Hamiltonian system with random initial conditions. The infinite dimensionality of the original Hamiltonian dynamics with random initial conditions (or, equivalently, the non-Markovianity of the finite dimensional stochastic dynamics (1)) renders the analysis of this dynamical system very difficult. This problem was studied in detail by Jaksic and Pillet in a series of papers [20, 22, 21]. In these works, existence and uniqueness of solutions as well as the ergodic properties of (1) were studied in detail. In particular, it was shown that the process $\{q, p = \dot{q}\}$ is mixing with respect to the measure

$$\mu_\beta(dqdp) = \frac{1}{Z_\beta} e^{-\beta H_{DP}(q,p)} \, dqdp.$$

To our knowledge, no information concerning the rate of convergence to equilibrium for the non-Markovian dynamics (1) is known for general classes of memory kernels. Ergodic theory for a quite general class of non-Markovian processes has been developed recently, see [12] and the references therein.

A class of memory kernels for which more detailed information on the long time asymptotics of the GLE (1) can be obtained was considered by Eckmann,

Hairer, Pillet and Rey-Bellet in a series of papers [37, 6, 7, 5]. Based on a generalization of Doob's theorem on stationary, Markovian, Gaussian processes [4], it was observed in these works that when the memory kernel $\gamma(t)$ has a rational spectral density, then the GLE (1) is equivalent to a **finite dimensional** Markovian system. This system is obtained by adding a finite number of additional degrees of freedom which account for the memory in the system. These auxiliary variables satisfy linear stochastic differential equations. As an example we mention the case where $\widehat{\rho}(k)$ in (4) can be written as

$$|\widehat{\rho}(k)|^2 = \frac{1}{|p(k)|^2}$$

where $p(k) = \sum_{m=1}^{M} c_m(-k)^m$ is a polynomial with real coefficients and roots in the upper half plane. Then the Gaussian process with spectral density $|\widehat{\rho}(k)|^2$ is the solution of the SDE

$$p\left(-i\frac{d}{dt}x(t)\right) = \frac{dW}{dt},$$

where $W(t)$ is a standard one dimensional Brownian motion. A related finite dimensional approximation of the infinite dimensional dynamics (1) has been introduced by Mori [30], see also [11] and the references therein. Mori's technique is based on a continued fraction expansion of the Laplace transform of the memory function $\gamma(t)$.

Motivated by the above, in this paper we will consider finite dimensional approximations of the GLE. The general form of the Markovian approximation of (1) can be written as [24] [1]

$$
\begin{aligned}
\dot{Q}_m(t) &= P_m(t), \quad Q_m(0) = q(0), &\text{(5a)}\\
\dot{P}_m(t) &= -\nabla V(Q_m(t)) + \lambda^T z(t), \quad P_m(0) = p(0), &\text{(5b)}\\
\dot{z}(t) &= -P_m(t)\lambda - Az(t) + C\dot{W}, \quad z(0) \sim \mathcal{N}(0, I), &\text{(5c)}
\end{aligned}
$$

where $z : \mathbb{R}^+ \mapsto \mathbb{R}^m$ and $\lambda \in \mathbb{R}^m$, $A, C \in \mathbb{R}^{m \times m}$. The fluctuation-dissipation theorem, which takes the form $CC^T = \beta^{-1}(A + A^T)$, is assumed to be satisfied.

In this paper we will consider (5) with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ and A diagonal with $A_{ii} = \alpha_i > 0$. This amounts to approximating the memory kernel by a sum of exponentials,

$$\gamma_m(t) = \sum_{i=1}^{m} \lambda_i^2 e^{-\alpha_i |t|}. \tag{6}$$

It is expected that the results proved in this paper are also valid in the more general case (5). As remarked in [24], when A is invertible, the more standard Mori approximation [30] is equivalent to (6) after an appropriate orthogonal transformation.

For this particular choice of λ and A the SDEs (5) become (we drop the

---

[1]To simplify the notation we consider (1) in one dimension. The results presented in this paper are valid in arbitrary finite dimensions. More details on the notation and on the multidimensional case can be found in Section 2.1 and Remark 3.3.

subscripts m for notational simplicity)

$$\dot{Q}(t) = P(t), \quad Q(0) = q(0), \tag{7a}$$

$$\dot{P}(t) = -\nabla V(Q(t)) + \sum_{j=1}^{m} \lambda_j z(t), \quad P(0) = p(0), \tag{7b}$$

$$\dot{z}_j(t) = -\lambda_j P(t) - \alpha_j z(t) + \sqrt{2\alpha_j \beta^{-1}} \dot{W}_j, \quad z_j(0) \sim \mathcal{N}(0, \beta^{-1}) \tag{7c}$$

for $j = 1, \ldots, m$. The process $\{Q(t), P(t), z(t)\}$ is Markovian with generator $-\mathcal{L}$ given by

$$-\mathcal{L} = p \cdot \nabla_q - \nabla_q V(q) \cdot \nabla_p + \sum_{j=1}^{m} \lambda_j z(t) \cdot \nabla_p$$

$$+ \sum_{j=1}^{m} \lambda_j \left(-\lambda_j p \nabla_{z_j} - \alpha_j z_j \cdot \nabla_{z_j} + \alpha_j \beta^{-1} \Delta_{z_j} \right). \tag{8}$$

This is a degenerate second order elliptic differential operator of hypoelliptic type [19]. Convergence to equilibrium for models of the form (7) has been studied using functional analytic techniques [7, 5]. Similar results have also been proved using Markov chain techniques [28, 37]. In this paper we present an alternative proof of exponentially fast convergence to equilibrium in relative entropy using the recently developed theory of hypocoercivity [40].

Several other results are also proved in this paper. We obtain sharp estimates on the derivatives of the Markov semigroup generated by $-\mathcal{L}$. We prove a homogenization theorem (invariance principle) for (7) in a periodic potential and we obtain estimates on the diffusion coefficient. Finally, we study the white noise limit of the GLE (1), i.e. the limit as the noise $F(t)$ in (1) converges to a white noise process. We show that in this limit the solution of (7) converges strongly to the solution of the Langevin equation

$$\ddot{q} = -V'(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}} \dot{W} \tag{9}$$

and we obtain a formula for the friction coefficient γ in terms of the coefficients $\{\lambda_j, \alpha_j\}_{j=1}^{m}$.

The rest of the paper is organized as follows. In Section 2 we state our main results and we introduce the notation that we will be using. In Section 3 we prove exponentially fast convergence to equilibrium in relative entropy. In Section 4 we prove estimates on the derivatives of the Markov Semigroup generated by $-\mathcal{L}$. In Section 5 we prove the homogenization theorem. In Section 6 we study the white noise limit. For the reader's convenience, background material on the theory of hypocoercivity is summarized in Appendix A. Finally, the proof of geometric ergodicity of the process (7) using Markov chain techniques is presented in Appendix B.

2 Statement of Main Results

We will use the notation $X := \mathbb{T}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$ and $Y := \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$. We will also denote the process $\{q(t), p(t), \mathbf{z}(t)\}$ by $\mathbf{x}(t)$. When we study the dynamics (7) in X the potential $V(q)$ is periodic, whereas when $\mathbf{x}(t) \in Y$ the

potential will be taken to be confining. The precise assumptions on the potential are given in Assumption 2.1 below.

Our fist result concerns the ergodicity of the SDE (7) in X or in Y. To prove the ergodicity of the SDE in Y we need to make the following assumptions on the potential.

Assumption 2.1.

(i) $V(q) \in C^2(\mathbb{R}^d)$ is a confining potential.

(ii) $\langle \nabla_q V, q \rangle \geq \sigma V(q) + \beta \|q\|^2$ for some suitable $\beta, \sigma > 0$ where $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the Euclidean inner product and norm, respectively.

(iii) There exists a constant c such that $\|\nabla^2 V\| \leq c$, where $\| \cdot \|$ denotes the Frobenious-Perron matrix norm.

The invariant measure $\mu_\beta(dq\, dp\, d\mathbf{z}) = \rho(p, q, \mathbf{z})\, dpdqd\mathbf{z}$ of the process (7), whose density satisfies the stationary Fokker-Planck equation $\mathcal{L}^* \rho = 0$, is known:

$$\rho(q, p, \mathbf{z}) = \frac{1}{Z} e^{-\beta(\frac{1}{2}|p|^2 + V(q) + \frac{1}{2}\|\mathbf{z}\|^2)}, \tag{10}$$

where $\mathcal{Z}$ is the normalization constant. This invariant measure is unique and the law of the process (7) converges exponentially fast to μ_β.

Theorem 2.1 (Ergodicity). *The process (7) with $\mathbf{x}(t) \in X$ and $V(q) \in C^3(\mathbb{T}^d)$ is geometrically ergodic. The same holds true when $\mathbf{x}(t) \in Y$, provided that the potential $V(q) \in C^3(\mathbb{R}^d)$ satisfies Assumption 2.1.*

The proof of this theorem, which is based on Markov chain-type arguments and which is similar to the proof presented in [37], see also [28], can be found in Appendix B.

We can prove exponentially fast convergence to equilibrium using tools from the theory of hypocoercivity [40]. We will use the notation $\mathcal{K} := \text{Ker}(\mathcal{L})$ and H^1_ρ for the weighted Sobolev space H^1 with respect to μ_β on either X or Y.

Theorem 2.2. *Let $-\mathcal{L}$ be the generator of the process $\mathbf{x}(t) \in X$, the solution of (7). Then there exist constants $C, \lambda > 0$ such that*

$$\|e^{-t\mathcal{L}}\|_{H^1_\rho / \mathcal{K} \to H^1_\rho / \mathcal{K}} \leq C e^{-\lambda t}.$$

The same holds true when $\mathbf{x}(t) \in Y$, provided that the potential satisfies Assumption 2.1(i) and (iii).

Using the tools from [40] we can prove exponentially fast convergence to equilibrium in relative entropy. The relative entropy (or Kullback Information) between two probability measures μ and ν with smooth densities f and ρ, respectively, is defined as

$$H_\rho(f) = \int f \log\left(\frac{f}{\rho}\right) d\mathbf{x}.$$

We will measure the distance between the law of the process $\mathbf{x}(t)$ at time t and the equilibrium distribution in relative entropy.

Theorem 2.3 (Convergence to Equilibrium). *Let f_t be the law of the process $\mathbf{x}(t)$ at time t and assume that $H_\rho(f_0) < +\infty$ and $V(q) \in C^2(\mathbb{T}^d)$. Then there exist constants C, $\alpha > 0$ such that*

$$H_\rho(f_t) \leq Ce^{-\alpha t}H_\rho(f_0).$$

The same holds true when $\mathbf{x}(t) \in Y$, assuming that $H_\rho(f_0) < +\infty$ and provided that the potential $V(q)$ satisfies Assumption 2.1(i) and (iii).

Remark 2.1. *In view of the Kullback inequality*

$$\frac{1}{2}\|f_t - \rho\|_{L^1} \leq H_\rho(f_t), \tag{11}$$

Theorem 2.3 implies that, for initial data with finite relative entropy, we have exponentially fast convergence to equilibrium in L^1.

The proofs of Theorems 2.2 and 2.3 are presented in Section 3.

Estimates on the Markov semigroup and its derivatives associated to the Langevin equation can be proved using an appropriate Lyapunov function [18, 13]. In this paper we use similar techniques to obtain estimates on the Markov semigroup and its derivatives for the generalized Langevin equation, equation (7). We introduce C_k, $k = 0, 1, 2$ with $C_0 = A$, $C_1 = [A, B]$ and $C_2 = [C_1, B]$. We will use the notation $L_\rho^2 := L^2(\,\cdot\,; \mu_\beta(d\mathbf{x}))$ where $\cdot$ is either X or Y.

Theorem 2.4 (Estimates on Derivatives of the Markov Semigroup). *Let $-\mathcal{L}$ be the generator of the process $\mathbf{x}(t) \in X$, the solution of (7) with $V(q) \in C^2(\mathbb{T}^d)$. Then the Markov semigroup $e^{-t\mathcal{L}}$ satisfies the bounds*

$$\|C_k e^{-t\mathcal{L}}\|_{L_\rho^2 \to L_\rho^2} \leq C\frac{1}{t^{\frac{1+2k}{2}}}, \quad k = 0, 1, 2 \text{ and } t \in (0, 1]. \tag{12}$$

The same holds true when $\mathbf{x}(t) \in Y$, provided that the potential $V(q)$ satisfies Assumption 2.1(i) and (iii).

Remark 2.2. *This result can also be obtained by applying Theorem A.3. Malliavin calculus-type arguments show that estimate (12) is sharp.*

When the potential $V(q)$ is periodic, the particle position, appropriately rescaled, converges weakly to a Brownian motion with a diffusion coefficient which can be calculated in terms of the solution of an appropriate Poisson equation. Results of this form have been known for a long time for the Smoluchowski (overdamped) equation [34, Ch. 13] as well as for the Langevin dynamics [33, 14]. In this paper we prove a similar result for the generalized Langevin equation. We will use the notation $\phi^e := \phi \cdot e$, $p^e := p \cdot e$, where e denotes an arbitrary unit vector in $\mathbb{R}^d$.

Theorem 2.5 (Homogenization). *Let $\mathbf{x}(t) \in X$ be the solution of (7) with $V(q) \in C^\infty(\mathbb{T}^d)$ with stationary initial conditions. Then the rescaled process $q_\epsilon^e(t) := e \cdot \epsilon q(t/\epsilon^2)$ converges weakly on $C([0, T], \mathbb{R})$ to a Brownian motion with diffusion coefficient D with*

$$D^e := De \cdot e = \beta^{-1} \sum_{j=1}^{m} \alpha_j \|\partial_{z_j}\phi^e\|^2, \tag{13}$$

where $\phi^e \in L_\rho^2$ is the unique, smooth, mean zero, periodic in q solution of the Poisson equation

$$\mathcal{L}\phi^e = p^e \tag{14}$$

on X. Furthermore, the following estimates hold

$$0 < D^e \leq \frac{4}{\beta} \sum_{i=1}^{m} \frac{\alpha_i}{\lambda_i^2}. \tag{15}$$

Let $q(t)$ be the solution of the Langevin equation (9) and let $q^\gamma(t) := q(\gamma t)$. It is well known that this rescaled process converges in the overdamped limit $\gamma \to +\infty$ to the solution of the Smoluchowski equation [31, Ch. 10]

$$\dot{q} = -\nabla V(q) + \sqrt{2\beta^{-1}}\dot{W}. \tag{16}$$

Similar results have also been proved in infinite dimensions [38]. In this paper we prove a similar result for the convergence of solutions to the GLE to the Langevin equation in the strong topology and obtain a formula for the friction coefficient that appears in the Langevin equation.

Consider (1) with the rescaled noise process

$$F^\epsilon(t) := \frac{1}{\sqrt{\epsilon}} F(t/\epsilon), \tag{17}$$

which is a mean zero stationary Gaussian process with autocorrelation function

$$\gamma^\epsilon(t) = \frac{1}{\epsilon}\gamma(t/\epsilon). \tag{18}$$

For the memory kernel (6), $\gamma^\epsilon(t)$ becomes

$$\gamma^\epsilon(t) = \sum_{j=1}^{m} \frac{\lambda_j^2}{\epsilon} e^{-\frac{\alpha_j}{\epsilon}|t|}. \tag{19}$$

Consequently, the rescaled noise process (17) is obtained by rescaling the coefficients in (7) according to $\lambda_j \to \frac{\lambda_j}{\sqrt{\epsilon}}$, $\alpha_j \to \frac{\alpha_j}{\epsilon}$. Under this rescaling the SDEs become

$$\begin{cases} \dot{q}(t) &= p(t), \\[2mm] \dot{p}(t) &= -\partial_q V(q) + \sum_{i=1}^{m} \frac{\lambda_i}{\sqrt{\epsilon}} z_i(t), \\[2mm] \dot{z}_i(t) &= -\frac{\lambda_i}{\sqrt{\epsilon}} p_t - \frac{\alpha_i}{\epsilon} z_i(t) + \sqrt{\frac{2\alpha_i\beta^{-1}}{\epsilon}}\dot{W}_i, \quad i = 1,...,m. \end{cases} \tag{20}$$

Theorem 2.6 (The White Noise Limit). *Let $\{q(t), p(t), \mathbf{z}(t)\} \in X$ be the solution of (20) with $V(q) \in C^1(\mathbb{T}^d)$ and initial conditions having finite moments of any order. Then $\{q(t), p(t)\}$ converge strongly to the solution of the Langevin equation*

$$\begin{cases} \dot{Q}(t) &= P(t), \\[2mm] \dot{P}(t) &= -\nabla_q V(Q_t) - \sum_{i=1}^{m} \left(\frac{\lambda_i^2}{\alpha_i} P(t) - \sqrt{\frac{2\lambda_i^2}{\alpha_i\beta^{-1}}}\dot{W}_i \right). \end{cases} \tag{21}$$

Consequently, the process $\{q, p\}$ converges weakly to the solution of the Langevin equation

$$\begin{cases} \dot{Q}(t) &= P(t), \\ \dot{P}(t) &= -\nabla_q V(Q_t) - \gamma P + \sqrt{2\gamma\beta^{-1}}\dot{W}, \end{cases} \tag{22}$$

where the friction coefficient γ is given by the formula

$$\gamma = \sum_{j=1}^{m} \frac{\lambda_i^2}{\alpha_i}.$$

2.1 Notation

For $\mathbf{x}(t) = (q, p, \mathbf{z}) \in Y := \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$ or $\mathbf{x}(t) \in X := \mathbb{T}^d \times \mathbb{R}^d \times \mathbb{R}^{dm}$ consider the operator $\mathcal{L}$:

$$\begin{aligned} -\mathcal{L} &= p\nabla_q - \nabla_q V(q)\nabla_p + \left(\sum_{j=1}^{m}\lambda_j z_j\right)\nabla_p \\ &\quad + \sum_{j=1}^{m}\left(-\alpha_j z_j \nabla_{z_j} - \lambda_j p\nabla_{z_j} + \beta^{-1}\alpha_j \nabla_{z_j}^2\right), \end{aligned} \tag{23}$$

with kernel $\mathcal{K} := Ker\mathcal{L}$. The density of the invariant measure of the process $\mathbf{x}(t)$ is

$$\rho(p, q, z) = \frac{1}{\mathcal{Z}}e^{-\beta(V(q)+\frac{1}{2}|p|^2+\frac{1}{2}|z|^2)}, \quad \mathcal{Z} = \int e^{-\beta(V(q)+\frac{1}{2}|p|^2+\frac{1}{2}|z|^2)}dpdqdz, \tag{24}$$

where $|\cdot|$ denotes either the Euclidean or the matrix norm. In (23), ∇ is the gradient (or the derivative when $d = 1$) and Δ the Laplacian. ∇^2 denotes the Hessian and if O is an operator then O^* is its adjoint in $L_\rho^2 := L^2(\cdot; \mu_\beta(d\mathbf{x}))$. Furthermore, we will use the notation $L = -\mathcal{L}$ so that L is the actual generator of the process.
Define

$$B = -p\nabla_q + \nabla_q V \nabla_p - \sum_{j=1}^{m}\lambda_j\left(z_j\nabla_p - p\nabla_{z_j}\right). \tag{25}$$

We easily check that $B^* = -B$. If $m = 1$ then $A_i = -\partial_{z_i}$ (derivative with respect to the $i-th$ component of z) so that $A_i^* = -z_i + \partial_{z_i}$ and we can write

$$\mathcal{L} = B + \sum_{i=1}^{d} A_i^* A_i = B + A^* A, \tag{26}$$

where A is intended to be the row vector of operators $(A_1, ..., A_d)$ (the same for A^*). More precisely, if $m = 1$ then: $A : L_\rho^2 \longrightarrow L_\rho^2 \otimes \mathbb{R}^d$, $B : L_\rho^2 \longrightarrow L_\rho^2$, $[A^*, A] : L_\rho^2 \longrightarrow L_\rho^2$, being $[A^*, A] := \sum_{j=1}^{d}[A_j^*, A_j]$; on the other hand $[A, A^*] : L_\rho^2 \longrightarrow L_\rho^2 \otimes \mathbb{R}^d \otimes \mathbb{R}^d$ is a matrix whose ij-th component is given by $[A, A^*]_{ij} := [A_i, A_j^*]$; in an analogous way $[A, A] : L_\rho^2 \longrightarrow L_\rho^2 \otimes \mathbb{R}^d \otimes \mathbb{R}^d$ is a

matrix with $[A, A]_{ij} := [A_i, A_j]$; finally $C := [A, B]$, $C : L_\rho^2 \longrightarrow L_\rho^2 \otimes \mathbb{R}^d$ is a vector of operators, $C_i = [A_i, B]$, $i = 1...d$, and the same holds for $C_2 := [C, B]$, $C_2 : L_\rho^2 \longrightarrow L_\rho^2 \otimes \mathbb{R}^d$.

If $m > 1$ then (26) becomes

$$\mathcal{L} = B + \sum_{i=1}^{m} \sum_{j=1}^{d} A_{ij}^* A_{ij} \tag{27}$$

with $A_{ij} = -\partial_{z_{i_j}}$ i.e. derivative with respect to the j-th component of z_i. We will use the notation

$$\mathcal{L} = B + A^* A, \tag{28}$$

meaning either (26) or (27).

As for the norms, unless otherwise specified, $\| \cdot \|$ indicates the norm of L_ρ^2, $\| \cdot \|_1^2 = \|A \cdot \|^2 + \|C \cdot \|^2 + \|C_2 \cdot \|^2$ is a sort of homogeneous $H^1(Y; \mu_\beta(d\mathbf{x}))$ norm and $\| \cdot \|_{H^1}^2 = \| \cdot \|^2 + \|A \cdot \|^2 + \|C \cdot \|^2 + \|C_2 \cdot \|^2$ is the usual inhomogeneous one. The inner products in these Hilbert spaces are denoted by $(\cdot, \cdot)$, $(\cdot, \cdot)_1$ and $(\cdot, \cdot)_{H^1}$, respectively.

Remark 2.3. *It is a general result that if $\mathcal{L}$ is accretive and $\bar{\mathcal{L}}$ is maximally accretive then $\bar{\mathcal{L}}$ is the generator of a contraction semigroup, so $e^{-\bar{\mathcal{L}}t}$ is well defined [17, Ch. 5].*
$\mathcal{L}$ is accretive since, as already noticed, $(\mathcal{L}f, f) = \|Af\|^2 \geq 0 \,\forall f$ in the domain of $\mathcal{L}$. To prove that $\bar{\mathcal{L}}$ is maximally accretive we show that the operator $\mathcal{L}^* + \nu I$ is injective for all $\nu > 0$. Since the operator is linear, we just need to show that $Ker(\mathcal{L}^* + \nu I) = \{0\}$ for all $\nu > 0$. So suppose $f \in Ker(\mathcal{L}^* + \nu I)$, then

$$\mathcal{L}^* f + \nu f = -Bf + A^* A f + \nu f = 0.$$

Consequently

$$(-Bf, f) + \|Af\|^2 + \nu \|f\|^2 = 0.$$

Therefore

$$0 \leq \|Af\|^2 = -\nu \|f\|^2 \Rightarrow f = 0.$$

Notice that when $\nu = 1$, from the above we conclude that $A^* A$ is self-adjoint and not only symmetric [35, Thm VIII.3]. From now on by $\mathcal{L}$ we mean its closure.

3 Convergence to Equilibrium

3.1 Hypocoercivity

Background material on hypocoercivity is presented in Appendix A.

Definition 3.1 (Hypocoercivity). *With the same notation as in Definition A.1 and assuming that the operator $\mathcal{T}$ generates a continuous semigroup, such an operator is said to be λ-hypocoercive on $\tilde{\mathcal{H}}$ if there exists a constant $\kappa > 0$ such that*

$$\| e^{-\mathcal{T}t} h \|_{\tilde{\mathcal{H}}} \leq \kappa e^{-\lambda t} \| h \|_{\tilde{\mathcal{H}}} \quad \forall h \in \tilde{\mathcal{H}} \text{ and } t \geq 0.$$

We say that an unbounded linear operator S on $\mathcal{H}$ is *relatively bounded* with respect to the (linear unbounded) operators $T_1, ..., T_n$ if $\mathcal{D}(S) \subset (\cap \mathcal{D}(T_j))$ and $\exists$ a constant $\alpha > 0$ s.t.

$$\forall h \in \mathcal{D}(S), \qquad \|Sh\| \leq \alpha(\|T_1 h\| + ... + \|T_n h\|).$$

The basic idea employed in the the theorems that we are going to use is to appropriately construct a scalar product on H_ρ^1 by adding lower order terms and then use the fact that hypocoercivity is invariant with respect to a change of equivalent norms, whereas coercivity does not enjoy such invariance. Finally, notice that $\mathcal{S}$, the class of Schwartz functions is dense in L_ρ^2, hence it is dense in $D(A) \cap D(B)$. This guarantees that all the operations performed on these (unbounded) operators are well defined.
Set $m = 1 = d$, $\alpha = \lambda = \beta = 1$. The first two commutators are

$$C_1 = C = [A, B] = \partial_p \quad \text{and} \quad C_2 = [C, B] = \partial_z - \partial_q. \tag{29}$$

Hence the operator is *hypoelliptic* [19]. Furthermore,

$$\begin{aligned}
& [A, A] = 0 \quad [A, C] = 0 \quad [A, C_2] = 0, \\
& [A, A^*] = Id \qquad [C, A^*] = 0 \qquad [C_2, A^*] = -Id, \\
& [C_2, B] = -\partial^2 V \partial_p - \partial_p, \\
& [C, C^*] = Id \qquad [C_2^*, C_2] = -Id - \partial_q^2 V,
\end{aligned} \tag{30}$$

where Id is the identity operator.

3.2 Proof of Theorem 2.2

Proof. We will use Theorem A.2 . To this end, set

$$P = A^* A + C^* C + C_2^* C_2$$

and notice that $Ker(P) = \mathcal{K} =: Ker\mathcal{L}$ contains only constants; in fact

$$Ker(P) = Ker(A^* A) \cap Ker(C^* C) \cap Ker(C_2^* C_2) = Ker(A) \cap Ker(C) \cap Ker(C_2).$$

To show that $\mathcal{K} = Ker(A^* A) \cap Ker(C^* C) \cap Ker(C_2^* C_2)$: the inclusion $\supseteq$ is obvious. For the other inclusion: if $h \in \mathcal{K}$ then $\|Ah\|^2 + \|Ch\|^2 + \|C_2 h\|^2 = 0 \Rightarrow Ah = Ch = C_2 h = 0$.
The above mentioned theorem requires two sets of hypotheses to be fulfilled. Hypothesis 1,2 and 3. in Theorem A.2 are quantitative assumptions, which are satisfied in our case with $N = 2$, $C_0 = A$, $C_1 = C$, $R_1 = R_2 = 0$, $R_3 = [C_2, B]$ (this is to have $C_3 = 0$) and thanks to Assumption 2.1(iii). Hypothesis 4. requires, in our case, for the operator P to be κ-*coercive* on $\mathcal{K}^\perp \cong L_\rho^2 / \mathcal{K}$. The coercivity of P is equivalent to

$$\|Ah\|^2 + \|Ch\|^2 + \|C_2 h\|^2 \geq \kappa \|h\|^2,$$

that is, more explicitly,

$$\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|(\nabla_z - \nabla_q) h\|^2 \geq \kappa \|h\|^2.$$

10

Using the fact that $\|a - b\|^2 \geq \frac{\|a\|^2}{3} - \frac{\|b\|^2}{2}$, we have

$$\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|(\nabla_z - \nabla_q) h\|^2 \geq \frac{1}{3} \left(\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|\nabla_q h\|^2 \right)$$

so we just need

$$\|\nabla_z h\|^2 + \|\nabla_p h\|^2 + \|\nabla_q h\|^2 \geq \kappa \|h\|^2$$

to hold true. Since μ_β is a product measure, we only need to verify that

$$\int |\nabla_q h|^2 e^{-V(q)} dq \geq \mu \int (h - \langle h \rangle)^2 e^{-V(q)} dq$$

holds true for some constant μ, where the notation $\langle h \rangle := \int h e^{-V(q)dq}$ has been used. It is a standard result that if $V(q) \in C^2(\mathbb{R}^d)$ is such that $e^{-V(q)}/\mathcal{Z}$ is a probability density and

$$\frac{|\nabla V(q)|^2}{2} - \Delta V(q) \overset{|q| \to \infty}{\longrightarrow} +\infty \tag{31}$$

then $e^{-V(q)}/\mathcal{Z}$ satisfies a Poincaré inequality (see, e.g., [40, Thm. A.1]). From Assumption 2.1(iii), Condition (31) is satisfied. We can conclude that there exist a scalar product $((\cdot, \cdot))$ inducing a norm equivalent to the inhomogeneous norm of H^1 and a constant $\hat{\lambda} > 0$ such that $\mathcal{L}$ is coercive in this norm:

$$\forall h \in L_\rho^2/\mathcal{K}, \qquad ((h, \mathcal{L}h)) \geq \hat{\lambda}((h, h)).$$

This implies that $\mathcal{L}$ is hypocoercive in this norm, hence it is hypocoercive on $L_\rho^2/\mathcal{K}$ endowed with the $\|\cdot\|_{H^1}$ norm:

$$\|e^{-t\mathcal{L}_{\gamma\beta}} h_0\|_{H^1} \leq C e^{-\lambda t} \|h_0\|_{H^1}. \tag{32}$$

$\square$

Remark 3.1. *The orthogonal space to $\mathcal{K}$ is the same with respect to both the $(\cdot, \cdot)_1$ and the $(\cdot, \cdot)_{H^1}$ norms; moreover, since P is coercive, these two norms are equivalent.*

Remark 3.2. *Theorem A.3 in Appendix A allows us to state a similar result when the initial datum is in L_ρ^2. In fact*

$$\|e^{-t\mathcal{L}} h\|_{H^1} \leq \frac{c}{t^{\frac{5}{2}}} \|h\|, \quad t \in (0, 1]. \tag{33}$$

So, putting together (32) and (33) we get, for $0 < t_0 < t$, $t_0 < 1$:

$$\begin{aligned}
\|e^{-t\mathcal{L}} h_0\|_{H^1} &= \|e^{-(t-t_0)\mathcal{L}} e^{-t_0\mathcal{L}} h_0\|_{H^1} = \|e^{-(t-t_0)\mathcal{L}} h_{t_0}\|_{H^1} \\
&\leq c e^{-\lambda(t-t_0)} \|h_{t_0}\|_{H^1} \leq c e^{-\lambda(t-t_0)} \|e^{-t_0\mathcal{L}} h_0\|_{H^1} \\
&\leq c \frac{e^{-\lambda(t-t_0)}}{t_0^{\frac{5}{2}}} \|h_0\|.
\end{aligned}$$

Remark 3.3. *The proof is identical when $m, d > 1$. In this case we can think of A as a matrix of operators, see (27).*

3.3   Proof of Theorem 2.3

Proof. Let f_t denote the law of the process $\mathbf{x}(t)$. We set $f_t = \rho h_t$. Then h_t satisfies the equation

$$\partial_t h_t = B h_t - A^* A h_t. \tag{34}$$

We apply Theorem A.4 to the operator $\mathcal{F} = -B + A^* A$ with

$$A = -\partial_z, \; C_1 = -\partial_p, \; C_2 = -\partial_q, \; Z_2 = Id, \; R_2 = -\partial_z.$$

Furthermore Asuumption 2.1 (i) and (iii) together with the Holley-Strook perturbation Lemma imply that $\mathcal{Z}^{-1} e^{-V(q)}$ satisfies a Logarithmic Sobolev Inequality (LSI).

Hypotheses 1,2 and 4 are automatically satisfied. We put $C_2 = \partial_q$ and we added the remainder R_2 in order to fulfill hypothesis 4. Hypothesis 3. is satisfied on account of Assumption 2.1 (iii). Now consider the relative entropy $H_\rho(f)$,

$$H_\rho(f) = \int f \log\left(\frac{f}{\rho}\right) dq \, dp \, dr = \int h \log h \, d\rho \qquad f = \rho h \tag{35}$$

and the Fisher information $I_\rho(f)$

$$I_\rho(f) = \int f |\nabla \log(h)|^2 dq dp dr = \int h |\nabla \log h|^2 \, d\rho \qquad f = \rho h. \tag{36}$$

Then if the initial datum has finite relative entropy, we obtain that

$$H_\rho(f_t) = \mathcal{O}(e^{-t\alpha}) \tag{37}$$

for some $\alpha > 0$ and for $t > 0$. If the initial datum has also finite Fisher information then

$$I_\rho(f_t) = \mathcal{O}(e^{-t\alpha}), \tag{38}$$

as well. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Remark 3.4. *We remark that (38), together with the LSI, implies (37).*

Remark 3.5. *In view of the LSI, it is interesting to notice that, applying Theorem A.5, we get the following bounds*

$$\int h(t) |C_k \log h(t)|^2 d\rho \le \frac{c}{t^{2k+1}} \int h_0 \log h_0 \, d\rho, \tag{39}$$

for $k = 0, 1, 2$ and c an explicitly computable positive constant.

4   Bounds on the derivatives of the Markov semigroup

Throughout this section we well be using the notation $u = e^{-t\mathcal{L}} u_0$. We introduce the Lyapunov function

$$\begin{aligned} F(t) \;=\; & a_0 t \|Au\|^2 + a_1 t^3 \|Cu\|^2 + a_2 t^5 \|C_2 u\|^2 \\ & + b_0 t^2 (Au, Cu) + t^4 b_1 (Cu, C_2 u) + b_2 \|u\|^2, \qquad t \in (0,1], \quad (40) \end{aligned}$$

where $a_j, b_j, j = 0, 1, 2$ are positive constants to be chosen.

Lemma 4.1. *There exist constants $a_j, b_j, j = 0, 1, 2$ such that the time derivative $\partial_t F$ of the Lyapunov function along the semigroup is negative.*

Proof. We will calculate the time derivative of each term in (40) separately and using the explicit relations (30):

$$
\begin{aligned}
\partial_t \|u\|^2 &= -2(\mathcal{L}u, u) = -2\|Au\|^2, \\
\partial_t(Au, Au) &= -2(Cu, Au) - 2\|A^*Au\|^2 = -2(Cu, Au) - 2\|Au\|^2 - 2\|A^2u\|^2, \\
\partial_t(Cu, Cu) &= -2\|ACu\|^2 - 2(C_2u, Cu), \\
\partial_t(C_2u, C_2u) &= ((2 + \partial_q^2 V)C_2u, Cu) - 2\|AC_2u\|^2 + 2(Au, C_2u), \\
\partial_t(Au, Cu) &= -2(A^2u, ACu) - (Au, Cu) - \|Cu\|^2 - (Au, C_2u), \\
\partial_t(Cu, C_2u) &= -\|C_2u\|^2 - 2(ACu, AC_2u) + 2\|Cu\|^2 + (Cu, Au).
\end{aligned}
$$

Putting everything together we obtain

$$
\begin{aligned}
\partial_t F(t) = {}& -2a_0 t\|A^2u\|^2 - 2a_1 t^3\|ACu\|^2 - 2a_2 t^5\|AC_2u\|^2 \\
& -2b_0 t^2 (A^2u, ACu) - 2b_1 t^4 (ACu, AC_2u) \\
& +(-2a_0 t + a_0 - 2b_2)\|Au\|^2 + (3a_1 t^2 + 2b_1 t^4 - b_0 t^2)\|Cu\|^2 \\
& +(5a_2 t^4 - b_1 t^4)\|C_2u\|^2 + (2b_0 t - 2a_0 t - b_0 t^2 + b_1 t^4)(Au, Cu) \\
& +(4b_1 t^3 - 2a_1 t^3 + 2a_2 t^5)(Cu, C_2u) + (2a_2 t^5 - b_0 t^2)(Au, C_2u).
\end{aligned}
$$

Now we estimate the sum of the first and of the second line (i.e. the sum of all the terms where A^2, AC and AC_2 appear). For $t \in (0, 1]$ we have

$$
(41)_1 + (41)_2 \leq -2a_0 t\|A^2u\|^2 + 2b_0 t^2\|A^2u\|\|ACu\|
$$

$$
+2b_1 t^4\|ACu\|\|AC_2u\| - 2a_1 t^3\|ACu\|^2 - 2a_2 t^5\|AC_2u\|^2
$$

$$
\leq -2a_0 t\|A^2u\|^2 + b_0^2 t\|A^2u\|^2 + t^3\|ACu\|^2 - 2a_1 t^3\|ACu\|^2
$$

$$
+b_1^2 t^3\|ACu\|^2 + t^5\|AC_2u\|^2 - 2a_2 t^5\|AC_2u\|^2.
$$

Similarly for the sum of the remaining terms (those with A, C and C_2) we have

$$
(41)_3 + (41)_4 + (41)_5 \leq (-2a_0 t + a_0 - 2b_2)\|Au\|^2
$$

$$
+(2b_0 t + 2a_0 t + b_0 t^2 + b_1 t^4)\|Au\|\|Cu\| + (2a_2 t^5 + b_0 t^2)\|Au\|\|C_2u\|
$$

$$
+(3a_1 t^2 + 2b_1 t^4 - b_0 t^2)\|Cu\|^2 + (5a_2 t^4 - b_1 t^4)\|C_2u\|^2
$$

$$
+(4b_1 t^3 + 2a_1 t^3 + 2a_2 t^5)\|Cu\|\|C_2u\|
$$

$$
\leq (-2a_0 t + a_0 - 2b_2)\|Au\|^2 + a_0^2\|Au\|^2 + \|Cu\|^2
$$

$$
+\frac{3}{2}b_0^2\|Au\|^2 + \frac{3}{2}t^2\|Cu\|^2 + \frac{1}{2}b_1^2\|Au\|^2 + \frac{t^4}{2}\|Cu\|^2
$$

$$
+a_2^2 t^5\|Au\|^2 + t^5\|C_2u\|^2 + \frac{t^2}{2}b_0^2\|Au\|^2 + \frac{t^2}{2}\|C_2u\|^2
$$

$$
+2b_1^2 t^3\|Cu\|^2 + t^3\|C_2u\|^2 + a_1^2 t^3\|Cu\|^2 + t^3\|C_2u\|^2
$$

$$+a_2^2 t^5 \|Cu\|^2 + t^5 \|C_2 u\|^2.$$

Choosing the constants in such a way that $\boxed{b_2 \gg a_0 \gg b_0 \gg a_1 \gg b_1 \gg a_2 > 1/c}$, where c is a constant depending on the bound on the second derivative of the potential, we obtain that $\partial_t F < 0 \ \forall t \in (0,1]$. $\square$

Proof of Theorem 2.4. We use the previous Lemma to deduce

$$a_0 t \|Au\|^2 + a_1 t^3 \|Cu\|^2 + a_2 t^5 \|C_2 u\|^2$$
$$+ b_0 t^2 (Au, Cu) + t^4 b_1 (Cu, C_2 u) + b_2 \|u\|^2 < b_2 \|u_0\|^2.$$

This, in turn, implies that

$$\|\nabla_z u\|^2 = \|Au\|^2 < \frac{\kappa}{t} \|u_0\|^2,$$

$$\|\nabla_p u\|^2 = \|Cu\|^2 < \frac{\kappa}{t^3} \|u_0\|^2,$$

$$\frac{\|\nabla_q u\|^2}{3} - \frac{\|\nabla_z u\|^2}{2} \leq \|\nabla_q u - \nabla_z u\|^2 = \|C_2 u\|^2 < \frac{\kappa}{t^5} \|u_0\|^2$$

$$\Rightarrow \|\nabla_q u\|^2 \leq \frac{\kappa}{t^5} \|u_0\|^2,$$

where κ is an explicitly computable positive constant. The previous inequalities are justified by the fact that

$$a_0 t \|Au\|^2 + a_1 t^3 \|Cu\|^2 + a_2 t^5 \|C_2 u\|^2 + b_0 t^2 (Au, Cu) + t^4 b_1 (Cu, C_2 u)$$

$$\geq (a_0 t - \frac{b_0^2}{2} t^2) \|Au\|^2 + (a_1 t^3 - \frac{t^2}{2} - t^4 \frac{b_1^2}{2}) \|Cu\|^2 + (a_2 t^5 - \frac{t^4}{2}) \|C_2 u\|^2$$

and the second line is positive thanks to the choice of the constants we made. $\square$

Remark 3.3 holds also in this case.

Remark 4.1. *From the estimates (12), similar estimates on $A^\star e^{-t\mathcal{L}^\bullet}$, $e^{-t\mathcal{L}^\star} A^\bullet$, $C^\star e^{-t\mathcal{L}^\bullet}$, $e^{-t\mathcal{L}^\star} C^\bullet$, $C_2^\star e^{-t\mathcal{L}^\bullet}$ and $e^{-t\mathcal{L}^\star} C_2^\bullet$ follow, where $\star$ and $\bullet$ stand for either the L_ρ^2-adjoint or nothing. In fact:*

(i) $(Ae^{-t\mathcal{L}} f, g) = (f, e^{-t\mathcal{L}^*} A^* g) \leq \|Ae^{-t\mathcal{L}} f\| \|g\| \leq \frac{\kappa}{\sqrt{t}} \|f\| \|g\|$
   $\Rightarrow (f, e^{-t\mathcal{L}^*} A^* g) \leq \frac{\kappa}{\sqrt{t}} \|f\| \|g\|$, *choose $f = e^{-t\mathcal{L}^*} A^* g$ and the result on $e^{-t\mathcal{L}^*} A^*$ follows.*

(ii) *Using $[A, A^*] = Id$ we have $\|A^* e^{-t\mathcal{L}} u_0\|^2 = \|A^* u\|^2 = \|Au\|^2 + \|u\|^2$, hence the estimate for $A^* e^{-t\mathcal{L}}$. Taking the adjoint as in (i) we get the result for $e^{-t\mathcal{L}^*} A$.*

(iii) *For $Ae^{-t\mathcal{L}^*}$ we can just repeat the proof we wrote for $Ae^{-t\mathcal{L}}$, since the only thing that changes when considering $\mathcal{L}^*$ is the sign of B, which doesn't play any role in the proof. Now, by acting as in (i) and (ii), we obtain the results for $e^{-t\mathcal{L}} A^*, A^* e^{-t\mathcal{L}^*}$ and $e^{-t\mathcal{L}} A$.*

14

5   The Homogenization Theorem

In this section we prove Theorem 2.5. The proof of this theorem is based on standard techniques, namely the central limit theorem for additive functionals of Markov processes [23, 27, 33], which in turn is based on the martingale central limit theorem [8, Thm. 7.1.4]. In order to apply these techniques we need to study the Poisson equation

$$\mathcal{L}u = f. \tag{41}$$

The boundary conditions for (41) are that $u \in L_\rho^2$ and it is periodic in q.

Proposition 5.1. *Let $f \in L_\rho^2 \cap C^\infty(X)$ with $\int_X f\mu_\beta(d\mathbf{x})$. Then the Poisson equation (41) has a unique smooth mean zero solution $u \in L_\rho^2 \cap C^\infty(X)$.*

The proof of Theorem 2.5 follows now from the above proposition.

Proof of Theorem 2.5. To simplify the notation we present the proof for $d = 1$. When $d > 1$ the same proof applies to the one-dimensional projections $q^e = q \cdot e$. In this case the diffusion coefficient D is replaced by the projections of the diffusion tensor $D^e = De \cdot e$.

We consider the process $\mathbf{x}(t)$ on X with stationary initial conditions. For non-stationary initial conditions we need to combine the analysis presented below with the exponential convergence to equilibrium, Theorem 2.2. Since $p \in L_\rho^2 \cap C^\infty(X)$ and centered with respect to the invariant measure $\mu_\beta(d\mathbf{x})$, Proposition 5.1 applies and there exists a unique mean zero solution $\phi \in L_\rho^2 \cap C^\infty(X)$ to the problem

$$\mathcal{L}\phi = p. \tag{42}$$

We use Itô's formula to obtain

$$d\phi = \mathcal{L}\phi\, dt + \sum_{j=1}^m \sqrt{2\alpha_j \beta^{-1}} \partial_{z_j} \phi\, dW_j.$$

We combine this, together with (42) and the equations of motion to deduce

$$
\begin{aligned}
q_\epsilon(t) &:= \epsilon q(t/\epsilon^2) \\
&= \epsilon q(0) + \epsilon \int_0^{t/\epsilon^2} p(s)\, ds \\
&= \epsilon q(0) - \epsilon\big(\phi(q(t/\epsilon^2), p(t/\epsilon^2), z(t/\epsilon^2) - \phi(q(0), p(0), z(0))\big) \\
&\quad + \epsilon \sum_{j=1}^m \int_0^{t/\epsilon^2} \sqrt{2\alpha_j \beta^{-1}} \partial_{z_j} \phi\, dW_j(s) \\
&=: \epsilon R^\epsilon + M^\epsilon.
\end{aligned}
$$

Our stationarity assumption, together with the fact that $\phi \in L_\rho^2$, imply that

$$\|R^\epsilon\| \le C.$$

To study the martingale term M^ϵ we use the martingale central limit theorem [8, Thm. 7.1.4] or [34, Thm 3.33]. We have that $M^\epsilon(0) = 0$, that $M^\epsilon(t)$ has continuous sample paths and, by stationarity, that it has stationary increments.

15

Furthermore, by stationarity and the fact that the Brownian motions $W_i(t)$, $i = 1, \ldots m$ are independent, we deduce that

$$\lim_{\epsilon \to 0} \langle M_t^\epsilon \rangle = 2 \sum_{i=1}^m \alpha_i \beta^{-1} \|\partial_{z_i} \phi\|^2 t \quad \text{in } L_\rho^1.$$

The above calculations imply that the rescaled one dimensional projection $q_\epsilon(t) := \epsilon q(t/\epsilon^2)$ converges weakly in $C([0,t]; \mathbb{R})$ to a Brownian motion $\sqrt{2D} W(t)$ where

$$D = 2 \sum_{i=1}^m \alpha_i \beta^{-1} \|\partial_{z_i} \phi\|^2. \tag{43}$$

Remark 5.1. *Notice that when $d > 1$ the convergence of the one dimensional projections does not imply the convergence of the process $q_\epsilon(t) = \epsilon q(t/\epsilon^2)$. The proof of this result, which is also based on the analysis of the Poisson equation, is very similar and it is omitted.*

To prove estimate (15), we first show the upper bound and than the fact that the diffusion coefficient is bounded away from zero (when we consider periodic solutions only). We set $\phi = g_i + \frac{1}{\lambda_i} z_i$ and use the Poisson equation (42) to obtain

$$\mathcal{L} g_i = -\frac{\alpha_i}{\lambda_i} z_i,$$

from which we obtain the estimate

$$
\begin{aligned}
\alpha_i \beta^{-1} \|\partial_{z_i} g_i\|^2 &\leq \sum_{j=1}^m \alpha_j \beta^{-1} \|\partial_{z_j} g_i\|^2 = (\mathcal{L} g_i, g_i) \\
&= \frac{\alpha_i}{\beta \lambda_i} \int g_i \partial_{z_i} \rho \, d\mathbf{x} = -\frac{\alpha_i}{\beta \lambda_i} \int \rho \, \partial_{z_i} g_i d\mathbf{x} \\
&\leq \frac{\alpha_i}{\beta \lambda_i} \|\partial_{z_i} g_i\|.
\end{aligned}
$$

Consequently $\|\partial_{z_i} g_i\| \leq \frac{1}{\lambda_i}$. From this we obtain the following estimate on the diffusion coefficient D

$$
\begin{aligned}
D &= \sum_{i=1}^m \alpha_i \beta^{-1} \|\partial_{z_i} \phi\|^2 = \frac{1}{\beta} \sum_{i=1}^m \alpha_i \|\partial_{z_i} g_i + \frac{1}{\lambda_i}\|^2 \\
&\leq \frac{2}{\beta} \sum_{i=1}^m \alpha_i \left(\|\partial_{z_i} g_i\|^2 + \frac{1}{\lambda_i^2} \right) \\
&\leq \frac{4}{\beta} \sum_{i=1}^m \frac{\alpha_i}{\lambda_i^2}.
\end{aligned}
$$

The fact that $D > 0$ is easily seen by contradiction. If $D = 0$ then, by (43), $\|\partial_{z_i} \phi\|^2 = 0 \; \forall i$. Hence $\phi = \phi(q, p)$ and

$$\mathcal{L} \phi = -p \partial_q \phi + \partial_q V \partial_p \phi + \sum_{i=1}^m \lambda_i z_i \partial_p \phi = p.$$

16

Multiplying both sides by $e^{z_i^2/2}$ and then integrating with respect to z_i we get

$$-\int p\partial_q\phi\, e^{z_i^2/2}dz_i + \int \partial_q V\partial_p\phi\, e^{z_i^2/2}dz_i$$
$$+\int \lambda_i z_i^2 e^{z_i^2/2}dz_i + \sum_{j\neq i}\int \lambda_i z_i z_j\partial_p\phi e^{z_i^2/2}dz_i$$
$$= \int pz\, e^{z_i^2/2}dz_i,$$

from which we conclude that $\lambda_i\partial_p\phi = 0$ for all i. Hence $\phi = \phi(q)$. By the same reasoning we get that $-p\partial_q\phi = p$, which does not have a periodic solution. $\square$

Proof of Proposition 5.1.
   The scheme of the proof, which is similar to the proof of [33, Lemma 2.1], is as follows. We consider the Poisson equation $\mathcal{L}\phi = f$ where $f \in L_\rho \cap C^\infty(X)$ and centered with respect to the invariant measure $\mu_\beta(d\mathbf{x})$.

1. We consider the modified problem

$$\lambda u_\nu + \mathcal{L}u_\nu + \nu\hat{P}u_\nu = f \qquad \lambda,\nu > 0 \tag{44}$$

   where $\hat{P}$ is a regularizing operator (i.e. the operator $\mathcal{L} + \hat{P}$ is uniformly elliptic) so that the Lax-Milgram theorem applies; hence the weak solution to (44) is unique $\forall \nu > 0$ and letting $\nu \to 0$, by the uniqueness of the weak limit, we get existence and uniqueness of the solution to

$$\lambda u + \mathcal{L}u = f \qquad \lambda > 0 \tag{45}$$

   (notice that $-\mathcal{L}$ is the generator of a Markov semigroup so by Hille-Yosida theorem the set $\{\lambda \in R : \lambda > 0\}$ is contained in the resolvent of the operator $-\mathcal{L}$)

2. Set $\mathcal{L}_\lambda u = \lambda u + \mathcal{L}u = \lambda u + f$ (the last equality has to hold in distribution if we want u to satisfy (41))

$$\Rightarrow u = \mathcal{L}_\lambda^{-1}(\lambda u + f) = \lambda\mathcal{L}_\lambda^{-1}u + \mathcal{L}_\lambda^{-1}f$$

   and defining $\mathcal{L}_\lambda^{-1}f = h$ we get $(\frac{1}{\lambda}Id - \mathcal{L}_\lambda^{-1})u = \tilde{h}$, $\tilde{h} = h/\lambda$.

3. Once proven that $\mathcal{L}_\lambda^{-1}$ is compact we are done; in fact in this case the Fredholm Theorem applies so either the solution to $(\frac{1}{\lambda}Id - \mathcal{L}_\lambda^{-1})u = \tilde{h}$ exists and is unique (and hence, by construction the solution to (41) is unique) or $(\frac{1}{\lambda}Id - \mathcal{L}_\lambda^{-1})u = 0$ admits a nonzero solution. We can rule out the latter option because $(\frac{1}{\lambda}Id - \mathcal{L}_\lambda^{-1})u = 0 \Leftrightarrow \mathcal{L}u = 0$; since we know that $Ker\mathcal{L}$ contains only constants and we are asking for u to have mean zero we can conclude that $\mathcal{L}u = 0 \Leftrightarrow u = 0$ and we are done.

Now the details. Choose $\hat{P} = C^*C + G^*G$ where $G = \nabla_q$ and notice that now $B = G^*C - C^*G + A^*C - C^*A$. Let us check that the hypothesis of Lax-Milgram Theorem hold:

$$\lambda(u_\nu, u_\nu) + \|Au_\nu\|^2 + \nu\|Cu_\nu\|^2 + \nu\|Gu_\nu\|^2$$

$$\geq \lambda(u_\nu, u_\nu) + \nu \|\nabla_{qpr} u_\nu\|^2 \geq min\{\lambda, \nu\} \|u_\nu\|_{H^1}$$

and

$$\lambda(u_\nu, v) + (Bu_\nu, v) + (Au_\nu, Av) + \nu(Cu_\nu, Cv) + \nu(Gu_\nu, Gv)$$
$$\leq \lambda(u_\nu, v) + \|Cu_\nu\|\|Gv\| + \|Gu_\nu\|\|Cv\| + \|Au_\nu\|\|Av\|$$
$$+\nu\|Cu_\nu\|\|Cv\| + \nu\|Gu_\nu\|\|Gv\| + \|Cu_\nu\|\|Av\| + \|Au_\nu\|\|Cv\|$$
$$\leq \alpha\|u_\nu\|_{H^1}\|v\|_{H^1},$$

with α a positive constant depending on ν and λ. Hence the weak solution to (44) exists and is unique. Moreover, from (44) we also have that

$$\lambda\|u_\nu\|^2 + \|Au_\nu\|^2 + \nu\left(\|Cu_\nu\|^2 + \|Gu_\nu\|^2\right) = (f, u_\nu) \leq \|f\|\|u_\nu\|,$$

hence $\lambda\|u_\nu\| \leq c$, $\|Au_\nu\|^2 \leq c$ and $\nu\|Cu_\nu\|^2, \nu\|Gu_\nu\|^2 \leq c$. Now taking u_ν/ν as test function we get that $\|u_\nu\|_{H^1} \leq c\|f\|$ and we can let $\nu \to 0$ obtaining that u_ν converges weakly to the solution of (45). Recalling that $u = \mathcal{L}_\lambda^{-1} f$ and that u the weak limit of $\{u_\nu\}$, we can write

$$\|\mathcal{L}_\lambda^{-1} f\|_{H^1} = \|u\|_{H^1} \leq \liminf_{\nu \to 0} \|u_\nu\|_{H^1} \leq c\|f\|$$

and using the fact that H_ρ^1 is compactly embedded in L_ρ^2 we can conclude that the resolvent is compact. This completes the proof. $\square$

6 The White Noise Limit

Throughout this section C denotes a generic constant, which is independent of ϵ. To simplify the notation we present the proof in one dimension. The proof is exactly the same in arbitrary dimensions. Let $(Q(t), P(t)) \in \mathbb{T} \times \mathbb{R}$ be the solution to the system (21). Then

$$| q(t) - Q(t) | \leq | q(0) - Q(0) | + \int_0^t | p(s) - P(s) | \, ds.$$

Setting $\theta_i = \lambda_i^2/\alpha_i$ we get that

$$\dot{p}(t) - \dot{P}(t) = -\partial_q V(q) + \partial_q V(Q) + \sum_{i=1}^m \theta_i(P(t) - p(t)) - \sqrt{\epsilon} \sum_{i=1}^m \frac{\lambda_i}{\alpha_i} \dot{z}_i(t).$$

Hence, for any $r > 2$,

$$\eta(T) := E \sup_{t \in [0,T]} \{| q(t) - Q(t) |^r + | p(t) - P(t) |^r\}$$

$$\leq CT^{r-1} \int_0^T E \sup_{s \in [0,t]} | q(s) - Q(s) |^r \, dt$$

$$+C\left(\sum_{i=1}^m \theta_i^r\right) T^{r-1} \int_0^T E \sup_{s \in [0,t]} | p(s) - P(s) |^r \, dt$$

$$+C\epsilon^{\frac{r}{2}} \sum_{i=1}^m \left(\frac{\lambda_i}{\alpha_i}\right)^r E \sup_{t \in [0,T]} | z_i(t) - z_i(0) |^r + E | p(0) - P(0) |^r + E | q(0) - Q(0) |^r.$$

18

From Gronwall's Lemma we get

$$\eta(T) \leq C \left[\sum_{i=1}^{m} \left(\frac{\lambda_i}{\alpha_i} \right)^r E \sup_{t \in [0,T]} \mid z_i(t) - z_i(0) \mid^r \right] \epsilon^{\frac{r}{2}} e^{CT}$$

and the result now follows from Proposition 6.1.

Proposition 6.1. *With the same notation and assumptions of Theorem 2.6 the following estimate holds true:*

$$\xi(T) \quad := \quad \sum_{i=1}^{m} E \sup_{t \in [0,T]} \mid z_i(t) \mid^r + E \sup_{t \in [0,T]} \mid p(t) \mid^r \leq C \left(1 + T + \frac{\alpha}{\epsilon} T + \sqrt{\frac{\alpha}{\epsilon}} T \right)$$

$$+ C \left(1 + \frac{\lambda}{\sqrt{\epsilon}} + \frac{\alpha + \sqrt{\alpha}}{\epsilon} \right) \int_0^T \xi(t) dt + E \mid p(0) \mid^r + \sum_{i=1}^{m} E \mid z_i(0) \mid^r .$$

In particular

$$\sum_{i=1}^{m} E \sup_{t \in [0,T]} \mid z_i(t) \mid^r \leq C \epsilon^{-1} (T + \sqrt{T}) e^{CT \epsilon^{\frac{r}{2} - 1}}$$

hence

$$\epsilon^{\frac{r}{2}} \sum_{i=1}^{m} \left(\frac{\lambda_i}{\alpha_i} \right)^r E \sup_{t \in [0,T]} \mid z_i(t) - z_i(0) \mid^r \longrightarrow 0 \qquad \forall r > 2.$$

Proof. We use Itô's formula to deduce that

$$\frac{1}{r} p(t)^r \quad = \quad \frac{1}{r} p(0)^r - \int_0^t (\partial_q V(q(s)) p(s)^{r-1}) \, ds + \frac{1}{\sqrt{\epsilon}} \sum_{i=1}^{m} \int_0^t \lambda_i p(s)^{r-1} z_i(s) \, ds$$

$$\frac{1}{r} z_i(t)^r \quad = \quad \frac{1}{r} z_i(0)^r - \frac{\alpha}{\epsilon} \int_0^t z_i(s)^r \, ds - \frac{\lambda}{\sqrt{\epsilon}} \int_0^t (p(s) z_i(s)^{r-1}) ds$$

$$+ \frac{\alpha}{\epsilon} \int_0^t z_i(s)^{r-2} ds + \sqrt{\frac{2\alpha_i}{\epsilon}} \int_0^t z_i(s)^{r-1} \, dW_i(s).$$

Setting $\lambda = \max \{\lambda_i, i = 1...m\}$, using the boundedness of $\partial_q V$ on the torus and Hölder's inequality, we have

$$E \sup_{t \in [0,T]} \mid p(t) \mid^r \leq CT + C \left(1 + \frac{1}{\sqrt{\epsilon}} \lambda \right) \int_0^T E \sup_{s \in [0,t]} \mid p(s) \mid^r \, dt$$

$$+ \frac{1}{\sqrt{\epsilon}} m \lambda \int_0^T \sum_{i=1}^{m} E \sup_{s \in [0,t]} \mid z_i(s) \mid^r \, dt + E \mid p(0) \mid^r .$$

Similarly,

$$\sum_{i=1}^{m} E \sup_{t \in [0,T]} | z_i(t) |^r \leq C \frac{1}{\sqrt{\epsilon}} \lambda \int_0^T E \sup_{s \in [0,t]} | p(s) |^r dt$$

$$+ C(\frac{\lambda}{\sqrt{\epsilon}} + 2\frac{\alpha}{\epsilon} + \sqrt{\frac{2\alpha}{\epsilon}}) \int_0^T \sum_{i=1}^{m} E \sup_{s \in [0,t]} | z_i(s) |^r dt$$

$$+ m\frac{\alpha}{\epsilon} T + Cm \sqrt{\frac{2T\alpha}{\epsilon}} + \sum_{i=1}^{m} E | z_i(0) |^r .$$

In the above we used the following estimate, which is a consequence of the Burkholder-Davies-Gundy inequality

$$E \sup_{t \in [0,T]} \left| \int_0^t z_i(s)^{r-1} dW_i(s) \right| \leq E \left| \int_0^T | z_i(s) |^{2(r-1)} ds \right|^{\frac{1}{2}}$$

$$\leq CE \left| \int_0^T (| z_i(s) |^{2r} + 1) ds \right|^{\frac{1}{2}}$$

$$\leq CE \sup_{t \in [0,T]} | z_i(t) |^r T^{\frac{1}{2}} + CT^{\frac{1}{2}}$$

$$\leq CT^{\frac{1}{2}} \int_0^T E \sup_{s \in [0,t]} | z_i(s) |^r dt + CT^{\frac{1}{2}}.$$

Putting everything together we obtain

$$\sum_{i=1}^{m} E \sup_{t \in [0,T]} | z_i(t) |^r + E \sup_{t \in [0,T]} | p(t) |^r$$

$$\leq C \left\{ T + m\frac{\alpha}{\epsilon} T + m\sqrt{\frac{2T\alpha}{\epsilon}} \right\}$$

$$+ C(1 + \frac{\lambda}{\sqrt{\epsilon}}) \int_0^T E \sup_{s \in [0,t]} | p(s) |^r dt$$

$$+ C(\frac{\lambda}{\sqrt{\epsilon}} + \frac{\alpha}{\epsilon} + \sqrt{\frac{\alpha}{\epsilon}}) \int_0^T \sum_{i=1}^{m} E \sup_{s \in [0,t]} | z_i(s) |^r dt$$

$$+ E | p(0) |^r + \sum_{i=1}^{m} E | z_i(0) |^r,$$

so that

$$\epsilon^{\frac{r}{2}} \xi(T) \leq C\epsilon^{\frac{r}{2}} + C\epsilon^{\frac{r}{2}-1}(T + \sqrt{T}) + C\epsilon^{\frac{r}{2}-1} \int_0^T ds\xi(s).$$

Applying Gronwall's Lemma gives then

$$\epsilon^{\frac{r}{2}} \xi(T) \leq C(T + \sqrt{T})\epsilon^{\frac{r}{2}-1} e^{CT\epsilon^{\frac{r}{2}-1}}$$

20

and, since $\epsilon^{\frac{r}{2}} \sum_{i=1}^{m} E \sup_{t \in [0,T]} \mid z_i(t) \mid^r \leq \epsilon^{\frac{r}{2}} \xi(T)$, we get the result. $\qquad \square$

A Hypocoercivity

In this appendix we recall some of the main results from the theory of hypocoercivity, as presented in [40]. Throughout this Appendix we will use the notation introduced in Section 2.1.

Definition A.1 (Coercivity). *Let $\mathcal{T}$ be an unbounded operator on a Hilbert space $\mathcal{H}$ with kernel $\mathcal{K}$. Let $\tilde{\mathcal{H}}$ be another Hilbert space continuously and densely embedded in $\mathcal{K}^\perp$. The operator $\mathcal{T}$ is said to be λ-coercive on $\tilde{\mathcal{H}}$ if*

$$(\mathcal{T}h, h)_{\tilde{\mathcal{H}}} \geq \lambda \|h\|_{\tilde{\mathcal{H}}}^2 \quad \forall h \in \mathcal{K}^\perp \cap D(\mathcal{T}).$$

The following Proposition gives an equivalent Definition of coercivity.

Proposition A.1. *With the same notation as in Definition A.1, $\mathcal{T}$ is λ-coercive on $\tilde{\mathcal{H}}$ iff*

$$\| e^{-\mathcal{T}t} h \|_{\tilde{\mathcal{H}}} \leq e^{-\lambda t} \| h \|_{\tilde{\mathcal{H}}} \quad \forall h \in \tilde{\mathcal{H}} \text{ and } t \geq 0.$$

Theorem A.2. *Let $\mathcal{L}$ be an operator of the form $\mathcal{L} = A^*A + B$, with $B^* = -B$, $\mathcal{K} = Ker\mathcal{L}$ and assume there exists $N \in \mathbb{N}$ such that*

$$[C_{j-1}, B] = C_j + R_j \quad 1 \leq j \leq N+1, \quad C_0 = A, C_{N+1} = 0. \qquad (46)$$

Consider the following assumptions: for $k = 0, ..., N+1$

1. *$[A, C_k]$ is relatively bounded with respect to $\{C_j\}_{0 \leq j \leq k}$ and $\{C_j A\}_{0 \leq j \leq k-1}$*

2. *$[C_k, A^*]$ is relatively bounded with respect I and $\{C_j\}_{0 \leq j \leq k}$ (here I indicates the identity operator on L_ρ^2)*

3. *R_k is relatively bounded with respect to $\{C_j\}_{0 \leq j \leq k-1}$ and $\{C_j A\}_{0 \leq j \leq k-1}$*

4. *$\sum_{j=0}^{N} C_j^* C_j$ is κ-coercive for some $\kappa > 0$*

If Assumptions $1 - 3$ are satisfied then there exists a scalar product $((\cdot, \cdot))$ on H^1 defining a norm equivalent to the usual H^1 norm and such that

$$\forall h \in H^1/\mathcal{K}, \quad ((h, \mathcal{L}h)) \geq K \sum_{j=0}^{N} \|C_j h\|^2 \qquad (47)$$

for some constant $K > 0$. Furthermore, if Assumption 4 is satisfied, then there exists a constant $\lambda > 0$ such that

$$\forall h \in H^1/\mathcal{K}, \quad ((h, \mathcal{L}h)) \geq \lambda((h, h)).$$

In particular, $\mathcal{L}$ is hypocoercive in $H^1/\mathcal{K}$, i.e.

$$\|e^{-t\mathcal{L}}\|_{H^1/\mathcal{K}\to H^1/\mathcal{K}} \le Ce^{-\lambda t}$$

for some $C, \lambda > 0$.

Theorem A.3. *With the same notation as in Theorem A.2, if Assumptions 1-3 are satisfied then*

$$\|C_k e^{-t\mathcal{L}}h\| \le \frac{C}{t^{k+\frac{1}{2}}}\|h\| \quad \forall k = 0, ..., N$$

for all functions $h \in L^2_\rho$.

Theorem A.4. *Let $V(x) \in C^2(\mathbb{R}^d)$ such that $\mu(dx) = e^{-V(x)}dx$ is a probability measure on $\mathbb{R}^d$ and assume that $\mathcal{L}$ generates a semigroup on a suitable space of positive functions. Let $\{A_j\}_{1 \le j \le M}$ and B be first order differential operators with smooth coefficients, with $B = -B^*$. Assume there exists $N \in \mathbb{N}$ such that*

$$[C_{j-1}, B] = C_j + R_j \quad 1 \le j \le N+1, \quad C_0 = A, C_{N+1} = 0.$$

If, for $0 \le k \le N+1$ the following assumptions are fulfilled

1. *$[A, C_k]$ is pointwise bounded with respect to A.*

2. *$[C_k, A^*]$ is pointwise bounded with respect to I and $\{C_j\}_{0 \le j \le k}$.*

3. *R_k is pointwise bounded with respect to $\{C_j\}_{0 \le j \le k-1}$.*

4. *$[A, C_k]^*$ is pointwise bounded relatively to I and A.*

5. *there is a positive constant $\lambda > 0$ such that $\sum_k C_k^* C_k \ge \lambda I$ pointwise on $\mathbb{R}^d$ (I is the identity matrix on $\mathbb{R}^d$).*

6. *The probability measure μ satisfies a logaritmic Sobolev inequality.*

Then the Kullback information and the Fisher information decay exponentially fast to zero.

Theorem A.5. *With the same notation as in Theorem A.4, let $V(x) \in C^2(\mathbb{R}^d)$ be such that $\mu(dx) = e^{-V(x)}dx$ is a probability measure on $\mathbb{R}^n$ and assume that $\mathcal{L}$ generates a semigroup on a suitable space of positive functions. If Assumptions $1 - 4$ of Theorem A.4 are fulfilled, then, along the semigroup, the following bounds hold*

$$\int h_t \mid C_k \log h_t \mid^2 d\mu \le \frac{C}{t^{2k+1}} \int h_0 \log h_0 d\mu \quad \forall k = 0, ..., N,$$

where $h_t = f_t/\rho$ and f_t is the law of the process with generator $L = -\mathcal{L}$.

B Ergodicity

In this appendix we apply Markov chain techniques [29, 28, 37] to prove ergodicity of the Markov process $\mathbf{x}(t) := \{q(t), p(t), \mathbf{z}(t)\}$ given by (7). To simplify the notation, we set all constants equal to 1. We will consider the ergodic properties of the SDE

$$\begin{cases} \dot{q} & = & p \\ \dot{p} & = & -\nabla_q V(q) + r \\ \dot{r} & = & -p - r + \dot{W}. \end{cases}$$

We study either the case $q \in \mathbb{R}^d$ or $q \in \mathbb{T}^d$ and $p, r \in \mathbb{R}^d$; we will sometimes use the notation $x = (q, p, r)$ and $L = -\mathcal{L}$ is the generator of the process.

Motivated by [28], consider the following conditions:

Lyapunov Condition : There exists a function $G(x) : R^{3d} \to [1, \infty)$ such that $G(x) \to \infty$ as $\|x\| \to \infty$ and $LG(x) \le -aG(x) + d$ for some $a, d > 0$.

Minorization condition : Let $P_t(x, A)$ be the transition kernel of the Markov process $\mathbf{x}(t)$. There exist $T > 0, \eta > 0$ and a probability measure ν, with $\nu(C^c) = 0$ and $\nu(C) = 1$ for some fixed compact set C in the phase space, such that

$$P_T(x, A) \ge \mu\nu(A) \quad \forall A \in \mathcal{B}(\mathbb{R}^{3d}), \, x \in C.$$

We will use the following result, whose proof can be found in [29].

Theorem B.1. : *Minorization condition + Lyapunov condition $\Rightarrow$ ergodicity.*

Assumption $(\star)$: Let $B_s(y) \in R^{3d}$ be the ball of radius s centered in y. For some fixed compact set C we have

- $P_t(x, A)$ has a density $p_t(x, y)$ which is continuous $\forall (x, y) \in C \times C$, more precisely

$$P_t(x, A) = \int_A p_t(x, y) \, dy \quad \forall A \in \mathcal{B}(\mathbb{R}^{3d}) \cap \mathcal{B}(C), \forall x \in C;$$

- $\forall x \in C$ and $\forall \delta > 0$ one can find a $\bar{t} = \bar{t}(\delta)$ such that

$$P_{\bar{t}}(x, B_\delta(x^*)) > 0 \quad \text{for some } x^* \in int(C), \forall x \in C.$$

Lemma B.1. *Assumption $(\star) \implies$ Minorization Condition.*

Now let $V(q)$ be any $C^1(\mathbb{T}^d)$ potential, $V(q) > -k$ for some positive constant k. Consider the function

$$G(x) = \hat{C} + \frac{B}{2}\|p\|^2 + \frac{C}{2}\|r\|^2 + DV(q) + H(p, r),$$

where B, C, D, H and $\hat{C}$ are positive constants to be chosen. We have that

$$G(x) \ge \hat{C} + \frac{B}{2}\|p\|^2 + \frac{C}{2}\|r\|^2 - \frac{H}{2}\|p\|^2 - \frac{H}{2}\|r\|^2 - Dk, \qquad (48)$$

so we need $B > H$, $C > H$ and $\hat{C} > Dk$. Moreover

$$
\begin{aligned}
LG(x) &= D(\nabla_q V, p) - B(\nabla_q V, p) - H(\nabla_q V, r) + B(r, p) + H\|r\|^2 \\
&\quad - C(p, r) - H\|p\|^2 - C\|r\|^2 - H(p, r) + C \\
&\leq H\|r\|^2 + \frac{H}{4}\|\nabla_q V\|^2 - H\|p\|^2 - C\|r\|^2 + C + H\|r\|^2,
\end{aligned}
$$

where we have chosen $B = D = C + H$. On the other hand, since $V(q) \leq K$,

$$
G(x) \geq -\frac{a}{2}B\|p\|^2 - \frac{a}{2}C\|r\|^2 - aKB - a\frac{H}{2}\|r\|^2 - a\frac{H}{2}\|p\|^2
$$

so imposing also $2H - C \leq -\frac{a}{2}(C + H)$, $-H \leq -\frac{a}{2}(B + H)$ for some $a > 0$, the Lyapunov condition is satisfied. One possible choice is $a = 1/4$, $B = 13/16$, $C = 5/8$ and H=3/16.
Notice that from what we have just proven it follows that $\forall l \geq 1$ we have

$$
LG(x)^l \leq -a_l G(x)^l + d_l, \tag{49}
$$

for some suitable positive constants a_l and d_l. In fact,

$$
\partial_{q_i} G(x)^l = lG(x)^{l-1}\partial_{q_i} G(x),
$$

$$
\partial_{p_i} G(x)^l = lG(x)^{l-1}\partial_{p_i} G(x),
$$

and

$$
\partial_{r_i}^2 G(x) = \partial_{r_i}\left[lG(x)^{l-1}\partial_{r_i} G(x)\right] = l(l-1)G(x)^{l-2}(\partial_{r_i} G)^2 + lG(x)^{l-1}\partial_{r_i}^2 G(x).
$$

Furthermore, using (48), we obtain

$$
l(l-1)G(x)^{l-2}(\partial_{r_i} G)^2 \leq c_l G(x)^{l-1},
$$

so that

$$
LG(x)^l \leq lG(x)^{l-1}LG(x) + c_l G(x)^{l-1}.
$$

Hence, using what we have proven in the case $l = 1$, we get (49).

Consider now the case $\mathbf{x}(t) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$. We introduce the Lyapunov function

$$
\begin{aligned}
G(x) &= \hat{C} + \frac{A}{2}\|q\|^2 + \frac{B}{2}\|p\|^2 + \frac{C}{2}\|r\|^2 + DV(q) \\
&\quad + E(p, q) + F(q, r) + H(p, r) + M(\nabla_q V, p).
\end{aligned}
$$

Consequently,

$$
\nabla_q G = Aq + D\nabla_q V + Ep + Fr + M\nabla^2 V(q) \cdot p,
$$

$$
\nabla_p G = Bp + Eq + Hr + M\nabla_q V,
$$

$$
\nabla_r G = Cr + Fq + Hp.
$$

24

Thus,

$$
\begin{aligned}
LG(x) \;=\; & A(p,q) + D(\nabla_q V, p) + E\|p\|^2 + F(p,r) - B(\nabla_q V, p) \\
& - E(\nabla_q V, q) - H(\nabla_q V, r) + B(p,r) + E(q,r) + H\|r\|^2 \\
& - C(p,r) - F(p,q) - H\|p\|^2 + M(p, Hess V(q) \cdot p) \\
& - C\|r\|^2 - F(r,q) - H(p,r) + C - M\|\nabla_q V\|^2 + M(r, \nabla_q V)
\end{aligned}
$$

From Assumption 2.1(iii) it follows that there exist constants $\tilde{\beta}$ and $\tilde{\sigma}$ such that

$$
\tilde{\sigma}\|q\|^2 - \tilde{\beta}\|\nabla_q V\|^2 \to +\infty \text{ as } \|q\|^2 \to +\infty.
$$

Hence, it follows that G satisfies the Lyapunov condition. Also in this case, one can prove that the Lyapunov condition holds for $G(x)^l$, $l \geq 1$, as well.

As for Assumption $(\star)$, first of all let us notice that, since the operator $\partial_t + \mathcal{L}$ is hypoelliptic, the transition probability has a density; because the SDE we are considering has time independent coefficients the density is C^∞ provided that $V(q) \in C^\infty$ [32]. Moreover, studying the control problem associated with $dx = b(x)dt + \sigma dw$, namely $dX = b(X)dt + \sigma dU$ where $U(t)$ is a smooth control, and using Strook-Varadhan support Theorem, we can prove that $P_t(x, A) > 0 \; \forall x \in \mathbb{R}^{3d}$, $t > 0$ and for any open $A \in \mathbb{R}^{3d}$.

Consider now the set $\mathcal{G}_l = \left\{ g : \mathbb{R}^{3d} \to \mathbb{R}, \text{ measurable} : \mid g(x) \mid \leq V(x)^l \right\}$. We have proven that the process $x(t)$ has a unique invariant distribution ρ. Furthermore, there exist constants $k = k(l)$ and $\lambda = \lambda(l)$, such that $\forall g \in \mathcal{G}_l$

$$
|E^{x_0} g(x(t)) - \rho(g)| \leq kG(x_0)e^{-\lambda t}, \quad t \geq 0. \tag{50}
$$

This completes the proof of the theorem.

References

[1] Ying SC. Ala-Nissila T, Ferrando R. Collective and single particle diffusion on surfaces. *Advances in Physics*, 51(3):949–1078, 2002.

[2] G. Ariel and E. Vanden-Eijnden. Testing transition state theory on Kac-Zwanzig model. *J. Stat. Phys.*, 126(1):43–73, 2007.

[3] G. Ariel and E. Vanden-Eijnden. A strong limit theorem in the Kac-Zwanzig model. *Nonlinearity*, 22(1):145–162, 2009.

[4] H. Dym and H. P. McKean. *Gaussian processes, function theory, and the inverse spectral problem*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. Probability and Mathematical Statistics, Vol. 31.

[5] J.-P. Eckmann and M. Hairer. Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Comm. Math. Phys.*, 212(1):105–164, 2000.

[6] J-P. Eckmann, C-A. Pillet, and L. Rey-Bellet. Entropy production in nonlinear, thermally driven Hamiltonian systems. *J. Statist. Phys.*, 95(1-2):305–331, 1999.

[7] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3):657–697, 1999.

[8] S.N. Ethier and T.G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[9] G. W. Ford, M. Kac, and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *J. Mathematical Phys.*, 6:504–515, 1965.

[10] D. Givon, R. Kupferman, and A.M. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):R55–R127, 2004.

[11] P. Grigolini. A generalized Langevin equation for dealing with nonadditive fluctuations. *J. Statist. Phys.*, 27(2):283–316, 1982.

[12] M. Hairer. Ergodic properties of a class of non-Markovian processes. In *Trends in stochastic analysis*, volume 353 of *London Math. Soc. Lecture Note Ser.*, pages 65–98. Cambridge Univ. Press, Cambridge, 2009.

[13] M. Hairer and G. A. Pavliotis. From ballistic to diffusive behavior in periodic potentials. *J. Stat. Phys.*, 131(1):175–202, 2008.

[14] M. Hairer and G.A. Pavliotis. Periodic homogenization for hypoelliptic diffusions. *J. Statist. Phys.*, 117(1-2):261–279, 2004.

[15] O. H. Hald and R. Kupferman. Asymptotic and numerical analyses for mechanical models of heat baths. *J. Statist. Phys.*, 106(5-6):1121–1184, 2002.

[16] P. Hanggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Modern Phys.*, 62(2):251–341, 1990.

[17] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.

[18] F. Hérau. Short and long time behavior of the Fokker-Planck equation in a confining potential and applications. *J. Funct. Anal.*, 244(1):95–118, 2007.

[19] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[20] V. Jakšić and C.-A. Pillet. Ergodic properties of the non-Markovian Langevin equation. *Lett. Math. Phys.*, 41(1):49–57, 1997.

[21] V. Jakšić and C-A. Pillet. Ergodic properties of classical dissipative systems. I. *Acta Math.*, 181(2):245–282, 1998.

[22] Vojkan Jakšić and Claude-Alain Pillet. Spectral theory of thermal relaxation. *J. Math. Phys.*, 38(4):1757–1780, 1997. Quantum problems in condensed matter physics.

[23] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.

[24] R. Kupferman. Fractional kinetics in Kac-Zwanzig heat bath models. *J. Statist. Phys.*, 114(1-2):291–326, 2004.

[25] R. Kupferman and A. M. Stuart. Fitting SDE models to nonlinear Kac-Zwanzig heat bath models. *Phys. D*, 199(3-4):279–316, 2004.

[26] R. Kupferman, A. M. Stuart, J. R. Terry, and P. F. Tupper. Long-term behaviour of large mechanical systems with random initial data. *Stoch. Dyn.*, 2(4):533–562, 2002.

[27] C. Landim. Central limit theorem for Markov processes. In *From classical to modern probability*, volume 54 of *Progr. Probab.*, pages 145–205. Birkhäuser, Basel, 2003.

[28] J.C. Mattingly and A. M. Stuart. Geometric ergodicity of some hypo-elliptic diffusions for particle motions. *Markov Processes and Related Fields*, 8(2):199–214, 2002.

[29] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[30] H. Mori. A continued-fraction representation of the time-correlation functions. *Progress of Theoretical Physics*, 34(3):399–416, 1965.

[31] E. Nelson. *Dynamical theories of Brownian motion.* Princeton University Press, Princeton, N.J., 1967.

[32] D. Nualart. *The Malliavin calculus and related topics.* Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[33] G.C. Papanicolaou and S. R. S. Varadhan. Ornstein-Uhlenbeck process in a random potential. *Comm. Pure Appl. Math.*, 38(6):819–834, 1985.

[34] G.A. Pavliotis and A.M. Stuart. *Multiscale methods*, volume 53 of *Texts in Applied Mathematics.* Springer, New York, 2008. Averaging and homogenization.

[35] M. Reed and B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis.* Academic Press, New York, 1980.

[36] L. Rey-Bellet. Open classical systems. In *Open quantum systems. II*, volume 1881 of *Lecture Notes in Math.*, pages 41–78. Springer, Berlin, 2006.

[37] L. Rey-Bellet and L. E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.*, 225(2):305–329, 2002.

[38] S. S. Cerrai and M. Freidlin. On the Smoluchowski-Kramers approximation for a system with an infinite number of degrees of freedom. *Probab. Theory Related Fields*, 135(3):363–394, 2006.

[39] I. Snook. *The Langevin and generalized Langevin approach to the dynamics of atomic, polymeric and colloidal systems.* Elsevier, Amsterdam, 2007.

[40] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[41] R. Zwanzig. Nonlinear generalized Langevin equations. *J. Stat. Phys.*, 9(3):215–220, 1973.